Serial No. 10/729,431
Wakefield, et al.
Filed December 5, 2003

Response to Office Action
Examiner: Phuong Thao Cao
Group Art Unit: 2184

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

INVENTORS: Todd D. Wakefield and David L. Bean

TITLE: **Methods and Products for Producing Role-Related
                   Information from Free Text Sources**

FILING DATE: December 5, 2003
                                        CONFIRMATION NO.: 8399
SERIAL NO.:   10/729,431

DOCKET: 5139 P

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

REMARKS

A terminal disclaimer is attached to overcome the double patenting rejection.

The claim objection has been addressed through the amendment above.

The section 101 rejection has been addressed through the amendment above.

The section 112 rejection has been addressed through the amendment above.

Applicant traverses the section 102 rejections. Gaizauskas pertains to extracting information from natural language text and recording that information in a structured representation (template). Gaizauskas does not, however, disclose the steps performed by Applicants' claimed invention. An overview of those steps is below, followed by a more detailed treatment of Gaizauskus:

accessing a source of data records that contain structured and

unstructured data, the unstructured data including free text, and the unstructured data of a particular record containing information related to the structured data in that record;

linguistically parsing the free text;

extracting multi-dimensional relational facts from the parsed free text, the multi-dimensional relational facts relating to the structured data of the data record from which the free text was taken, and the multi-dimensional relational facts including a plurality of attributes derived from the free text

said extracting step including identifying roles within the parsed text records, each of said extractions containing role information;

applying caseframes to extractions, said applying caseframes producing a filtered set of attribute extractions;

- and

producing a structured data element from the filtered set of attribute extractions, said filtered set of attribute extractions containing relational facts relating to said structured data.

In particular, Applicant has claimed accessing data records that contain BOTH structured and unstructured data. Gaizauskas accesses ONLY unstructured data. Applicant requests the Examiner to review the sections of Gaizauskas cited in the office action to confirm the fact that Gaizauskas does not access any data records that contain a combination of unstructured and structured data. Therefore Gaizauskas does not disclose this element found in Applicant's claims.

Applicant has also claimed the step of extracting relational facts from the free text contained in the data records that have both structured and unstructured data.

Response to Office Action
Examiner: Phuong Thao Cao
Group Art Unit: 2164

Gaizauskas discloses only use of unstructured data, therefore Gaisauskus does not disclose "extracting relational facts from the free text, said extracting step being performed using linguistic information from the free text" as claimed by Applicant.

Applicant has also claimed the step of producing construed data with relational facts from the unstructured data and then relating those facts to the structured data through an integration process.. Gaizauskus discloses only accessing unstructured data to fill his templates and perform his further work with the unstructured data alone. He does not produce relational facts from unstructured data and then relate them to structured data found in the same data record as the unstructured data. Therefore Gaizauskus does not disclose Applicant's integration step.

Applicant has also claimed the step of integrating the construed data with the structured data to create a useful result. Gaizauskus only accesses unstructured data and performs his work from there. He does not begin with data records that contain both structured and unstructured (free text) data mixed together, so he has no way of relating his results back to the structured data of the same data record that the free text came from. Consequently, Gaizauskus does not anticipate Applicant's claims.

Gaizauskus also does not disclose using linguistic information within, relying instead on formatting of, free text in order to create relational facts that relate the free text to structured data in the same data record in which the free text was found. As Gaizauskus does not use linquistic information in this manner, Gaizauskus does not anticipate Applicant's patent claims.

Therefore, Gaizauskus does not disclose the claimed invention for any of the pending

claims under section 102 of the Patent Act.

Further, Applicant wishes to explain the teachings of the Gaizauskus reference. The information extraction work described in Gaizauskus makes no reference to combining extracted information with structured data. From the Gaizauskus introduction, "IE may be seen as the activity of populating a structured information source (or database) from an unstructured, or free text, information source." Further, when Gaizauskus does discuss combining IE with another technology, it is with information retrieval (IR), another text-focused task that does not address structured data. Applicant provides a function of integrating relational facts from unstructured data back to the structured data in the same data record in which the unstructured data was found. Gaizauskus does not disclose this integration function at all. From the Gaizauskus introduction, "The contrast between the aims of IE and IR systems can be summed up as: IR retrieves relevant documents from collections, IE extracts relevant information from documents. The two techniques are therefore complementary, and their use in combination has the potential to create powerful new tools in text processing." Applicant's invention is not about combining IR and IE – it is about combining a particular style of IE with structured data to provide the foundation for a new type of data analysis, as recited in the claims.

The template-filling style of IE described in Gaizauskus is a targeted type of extraction, i.e. the template describes a set of predefined participants or attributes of an event to be populated by the extraction engine. For example, in a kidnapping event, the kidnapping template would have slots for the victim, the perpetrator, the location, and the time of the kidnapping. In one embodiment of Applicants' invention, Applicants describe a system that records all participants of all actions, in a generic form. So, for a kidnapping event, the invention would extract the actor and object of the kidnapping

action, but it would also extract the actors and objects of all other events in the text. These events might include physical descriptions of the victim, a list of previous activities by the perpetrator, and a historical recounting of terrorist activities in the area. Unless every possible participant in every possible event has been encoded in a set of templates of the prior art, the targeted approach that Gaizauskus discusses will not cover the exhaustive functions performed by Applicants Invention. The invention includes a style of IE that is fundamentally distinct from traditional IE technologies at work only in a targeted fashion.

With respect to specific statements about Gaizauskus made by the Examiner, Applicants believe that Gaizauskus does not make the full disclosure alleged. A comparison of Gaizauskus to the claims as performed above shows the novelty of Applicants' invention. Withdrawal of the rejection is requested.

Applicant also believes that the pending claims are allowable due to reciting use of caseframes. A caseframe is something can be applied to a parse tree, but it is not a parse tree not is it a script-driven extraction. Paragraph 12 of Applicant's patent application explains caseframes:

> "Caseframes, generally speaking, are patterns that identify a particular linguistic construction and an element of that construction to be extracted. A syntactic caseframe, for example, may be applied to a parsed sentence to identify a clause that contains a subject and an active voice verb, and to extract the subject noun phrase. A syntactic caseframe often also uses lexical filters to constrain its identification process. For example, a user might want to extract the names of litigation plaintiffs in legal documents by creating a caseframe that extracts the subjects of a single active voice verb, sue. Other caseframe types may be fashioned, such as thematic role caseframes that apply their patterns, not to syntactic constructions, but thematic role relationships. More than one caseframe may apply to a sentence. If desired, a selection process may be utilized to reduce the

number of caseframes that apply to a particular sentence, although under many cirumstances that will not desirable nor necessary."

Gaizauskas does not teach this concept. Thus all claims that recite use of caseframes are novel and nonobvious.

Applicant has clarified its claims through the amendment above that Applicant's invention uses <u>multi-dimensional relational facts</u>. Gaizauskas does not use multi-dimensional relational facts, relying instead on a simplistic script-driven process. Applicant's multi-dimensional relational facts include multiple attributes. For example, consider the following stentence:

"The bolt on the underside of the transmission case was cracked due to heat."

The present invention uses relational fact extraction that encompasses multiple dimensions – i.e., what failed = bolt; how did it fail = cracking; where it failed = underside of transmission case; why it failed = heat. In other words, the claimed invention uses relational fact extractions that include multiple potential dimensions or attributes going to who, what, when, where, why and how. This facilitates multi-dimensional analysis. The prior art does not accomplish this.
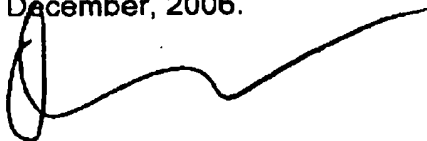
Applicant has claimed using records of mixed structured and unstructured data, using linquistic information to extract multi-dimensional relational fact from the unstructured data (free text), and then relating those relational facts to the structured data. Gaizauskas only accesses free text, rather than records which mix structured data with unstructured data (free text).

For these reasons, withdrawal of each of the rejections is appropriate. Reconsideration is requested.

Serial No. 10/729,431
Wakefield, et al.
Filed December 5, 2003

Response to Office Action
Examiner: Phuong Thao Cao
Group Art Unit: 2164

Respectfully submitted this 12th day of December, 2006.

Daniel P. McCarthy
Reg. No. 36,600
P.O. Box 71550
Salt Lake City, UT 84171-0550
Tel 801 661 8998
Fax 435 615 9669
Email: dmccart@xmission.com